

Introduction

The ImmPort system supports the National Institute of Allergy and Infectious Diseases (NIAID) and Division of Allergy, Immunology, and Transplantation (DAIT) mandate to facilitate storage, sharing and analysis of research data.

This tutorial is intended as an introduction to the ImmPort data submission process. It is meant to supplement, but not replace interaction with the ImmPort staff. Contact the ImmPort staff for the preparation of the data for submission and using the analytical tools at helpdesk@immport.org.

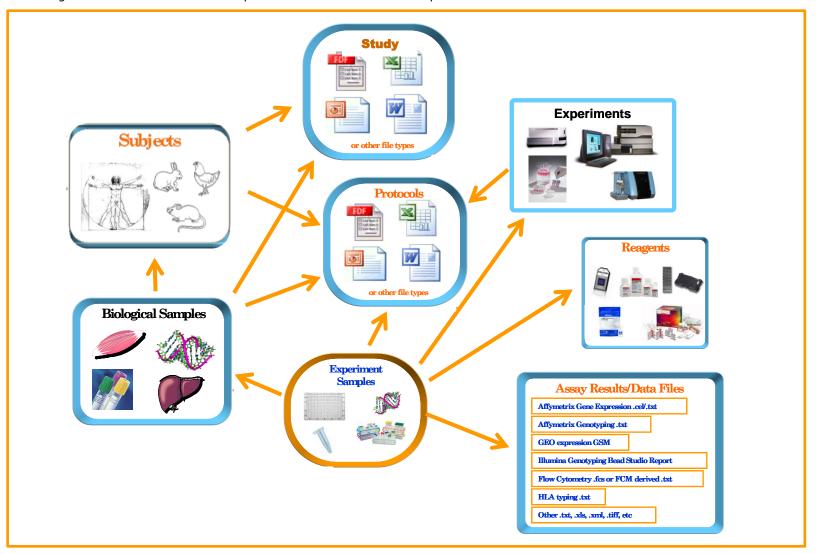
Data Submission allows you to upload descriptive information about Studies, Subjects, Samples, Experiments, Protocols, Reagents, and Results. An ImmPort data submission package may include a complete data set or incremental data sets. The packages are transferred by secure ftp to the ImmPort system. The contents of the package are processed to store the results and make them available for the analysis, query and sharing tools of ImmPort.

A Research Project generally refers to a private workspace to be used by an individual or by a group researchers collaborating on a given project. The workspace provides access control to studies, experiments, results, and any accompanying data. If you need assistance in understanding what a Research Project is and whether you are associated with one, please contact the Help Desk.



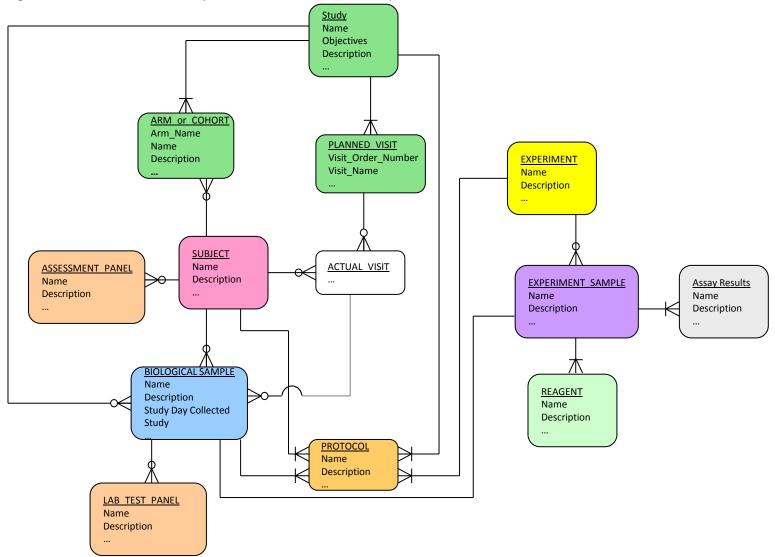
The ImmPort Data Model

This diagram illustrates the relationships between research data components.



Another View of the ImmPort Data Model

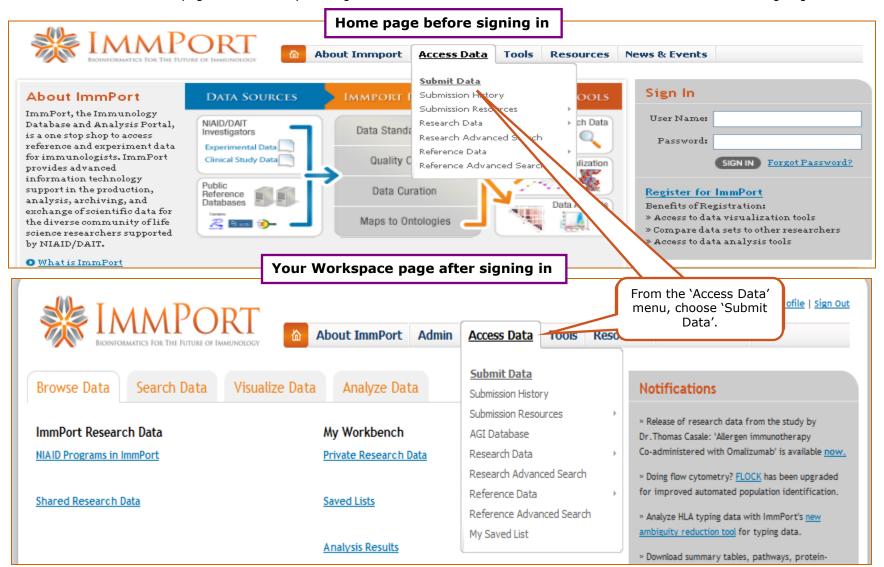
This diagram illustrates the relationships between research data components with more details





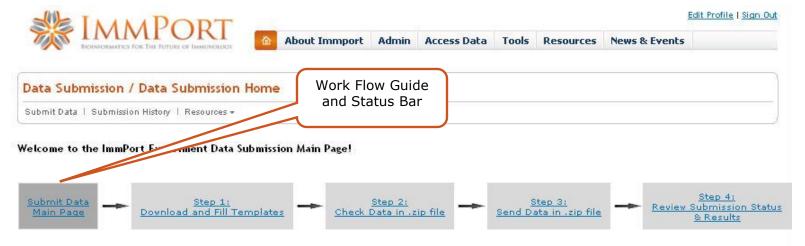
Data Submission

The Data Submission main page is accessed by clicking on the "Submit Data" link from the menu bar either before or after signing in.





Data Submission Main Page.



The Experiment Data Submission module allows users to upload experiment data into the ImmPort DATABASE. Information about the experiment and its results are organized into several components including Subjects, Biological Samples, Experiment Samples, Protocols, Reagents, and Experiment Sample Results.

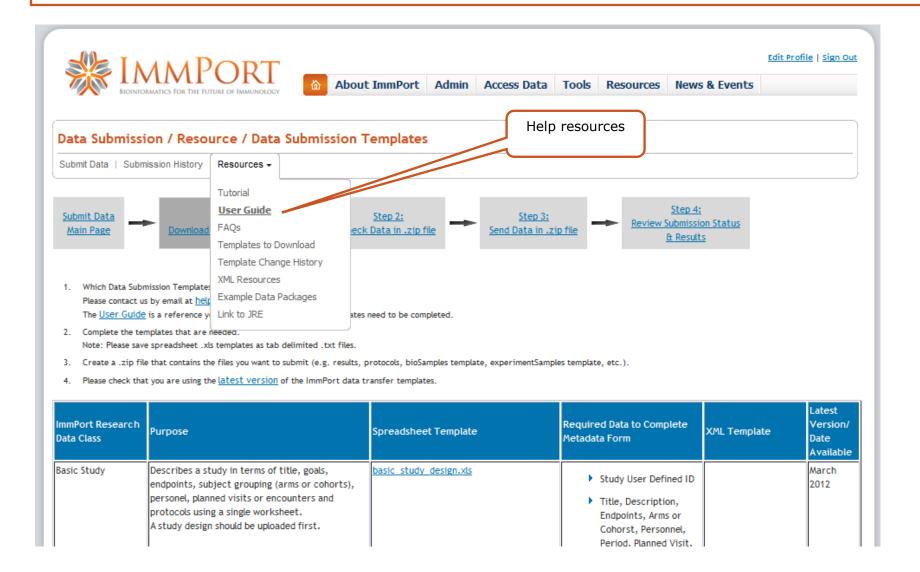
If this is your first time entering data, you will need the following:

- A set of protocol documents delineating your SOPs
- Reagent list
- A list of types of samples (blood vs lymphocytes vs serum, etc.)
- A system for uniquely defining each sample such that sample type, date of sample, type of analysis can be discerned

Please check that you are using the latest version of the ImmPort data transfer templates.

- Data Submission Template Change History
- User Guide
- ▶ Tutorial
- Example Data Packages
- Data Package Validator
 - This tool checks the format and content of the files in the data package and reports issues.
 - The tool is downloadable client side and executed from a graphical user interface.







the research.

RT ImmPort Data Submission

Metadata Transfer Templates

Step 1: Download and Complete Templates. The descriptive information about research data is captured in metadata transfer file templates. There are two formats provided on this site for the metadata templates: spreadsheet based and XML (eXtensible Markup Language) based. Both formats capture equivalent descriptive information.

There are 7 categories of metadata transfer templates basic study design- Describe the purpose and goals of

protocols.xls – CRFs, sample preparation, assay procedures

reagents.xls – Key components of assay (e.g. fluorochrome conjugated antibody)

subjectsHuman.xls – Human subjects information including demographics and phenotype

subjectsAnimal.xls – Non-human research subjects descriptions including strain and phenotype

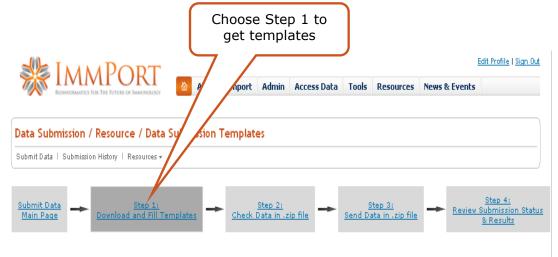
experiments.xls – Describe experiments, including the hypothesis and experimental variables

bioSamples.xls – Samples extracted and processed from subjects (e.g. blood, DNA) and applied treatments

experimentSamples.xls – samples, reagents, and results form an assay. Note that there are templates tailored for assay types.

There are additional templates to describe HLA and KIR typing systems

Derived or interpreted assay results are reported in a assay specific for ELISA, ELISPOT, Flow Cytometry, HAI, HLA, KIR, Virus Neutralization results.



- Which Data Submission Templates do you need?
 Please see the <u>User Guide</u> to determine which, if any templates need to be completed.
- Complete the templates that are needed.
 Note: Please save spreadsheet .xls templates as tab delimited .txt files
- Create a .zip file that contains the files you want to submit (e.g. results, protocols, bioSamples template, experimentSamples template, etc.).
- 4. Please check that you are using the <u>latest version</u> of the ImmPort data transfer templates.

ImmPort Research Data Class	Spreadsheet Template	Required Fields Checklist	XML Template	Latest Version/ Date Available
Protocols	protocols.xls	Protocol User Defined Id	protocols.xml	09/08/2008
Reagents	reagents.xls	Reagent User Defined Id	<u>reagents.xml</u>	09/08/2008
Subjects	subjects Human. x ls subjects Animal. x ls	Human User Defined Id	subjects.xml	09/08/2008
		▶ Protocol		



Data Submission Frequently Asked Questions

Which templates do I need?

A complete data submission package will include all of the completed templates and result files. Incremental data submission packages will include some of the completed templates and perhaps results.

Why are there so many templates?

ImmPort organizes descriptive information about research studies by commonly used domains or classes; studies, subjects, samples, protocols, experiments. The upload templates proves a means to capture relevant descriptive information about each data class and to link records from different data classes together.

Which templates should I complete first?

The protocol and reagent templates are good candidates to complete first as these are often available to the researcher at the beginning of a study and protocols and reagents can be referenced by other descriptive data types. Subjects and Biological Samples could be completed next when this data becomes available. Experiments and the samples associated with experiments

What sort of protocols should I describe?

Among the types of protocols to consider including in your data submission package are subject recruitment and evaluation forms, CRFs (Clinical Report Forms), sample extraction and treatment protocols and assay procedures.

Should I use Spreadsheet based or XML?

Both template formats capture equivalent data. Spreadsheets may be more intuitive to bench scientists and XML may be more applicable for informaticists.

Which columns do I need to complete?

There are a few required and several optional descriptive fields to complete in each template. To successfully load data into ImmPort, only the required fields in the templates need to be filled out. However, in order to fully use ImmPort's research data search capabilities, we encourage you to provide as much information as possible.

How do I know which columns are required?

In the spreadsheet templates, the required columns have a white background and the column name has a red asterisk. The column names have comments that can be seen by hovering the cursor over the column name.

What is a user defined ID?

The user defined ID is intended to be a short name for a metadata record (e.g a subject or sample) and must be unique within the scope of a research project.

How do I link data elements together?

The templates "talk" to each other via key linkages (for example, the Protocol IDs defined in protocols.xls will be referenced in bioSamples.xls when you indicate which protocol was used for a particular biological sample). It is important to make sure all of these key linkages match up when they are used.

Summary of links to and from the metadata transfer files

This table summarizes how data elements can be linked together.

Link To	Studies	Protocols	Reagents	Subjects	Biological Samples	Experiments	Experiment Samples	Results
Link from								
Studies		Yes						
Protocols		-	-	-	-	-	-	-
Reagents		-	-	-	-	-	-	-
Subjects	Yes (Study Arm or Cohort)	Yes	1	-	-	-	-	-
Biological Samples	Yes	Yes	-	Yes	-	-	-	-
Experiments		Yes	•	-	-	-	-	-
Experiment Samples		Yes	Yes	-	Yes	Yes	-	Yes
Results		-	-	-	-	-	-	-



Spreadsheet Metadata Transfer File Features

- Do not edit or delete the column headers. The column headers are used to ensure that the data the user enters is properly processed.
- •Non-shaded columns are **required fields** to be completed but the shaded columns are optional fields.
- •Explanatory **comments** which include the type of information and the size of text to enter into a cell are available by moving the cursor (mouse) over the column header name. Commented column headers have a small red triangle in the upper right corner of the cell.
- •**Drop down lists** for some columns (e.g. measurement technique in experiments.xls) are activated by clicking on a cell in the column (i.e. controlled vocabulary).
- •When working on the Excel templates, we recommend that you save these templates in Microsoft Excel format until you are ready to upload a submission package. At that point, you will save all templates as tab-delimited text file files.
- •In the spreadsheet templates, the following metadata categories use a **color code**:

Biological Samples: blue yellow Experiments: purple Protocols: orange Reagents: light green

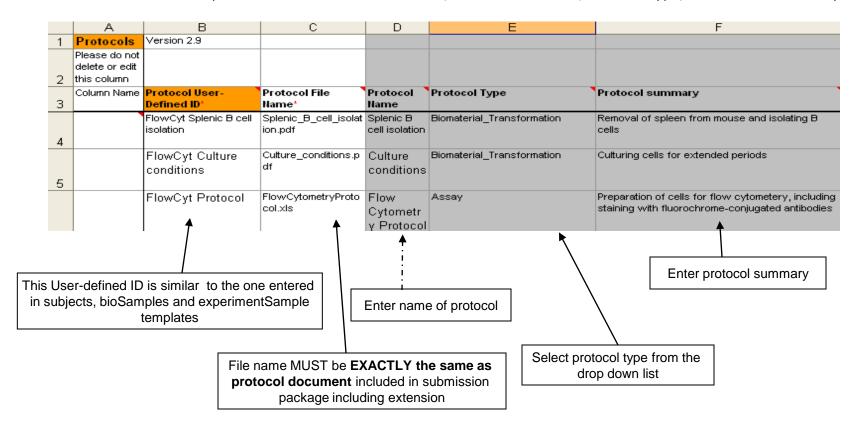
Subjects: rose

When linking entries in one metadata transfer file to another file, the color coding of column headers can be used as a guide to indicate how to make the link.

Protocols

The **protocols.xls** template captures information on how the Biological Sample was isolated, enriched, processed, treated or otherwise prepared for experimental use reagents to generate experiment results.

The columns in this template are Protocol User-Defined ID, Protocol File Name, Protocol Type, and Protocol summary.

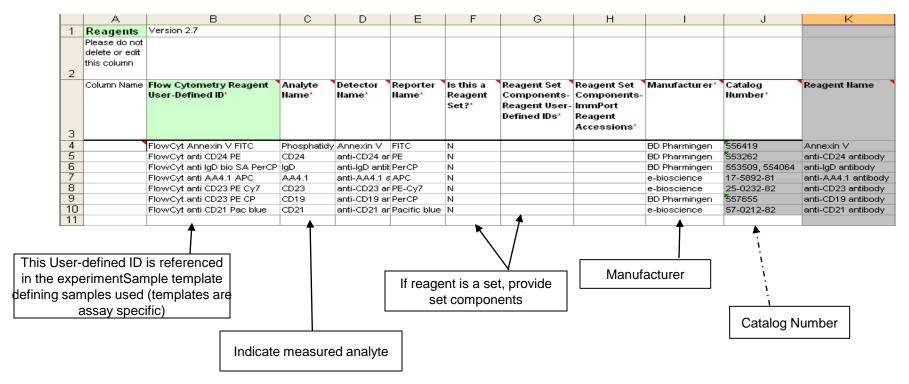




Reagents

The **reagents.xls** template captures chemical compounds used to measure an Analyte(s). Different unique files of reagents worksheets can be selected for Array, ELISA, ELISPOT, FCM, HLA, Illumina and Other. The example given is a completed reagents version of ELISPOT. Please refer to the Data Submission user guide for detailed information.

Whenever there is a comma in a text in any cell, Microsoft Excel will add a double quote at the beginning of the text and a double quote at the end of the text.

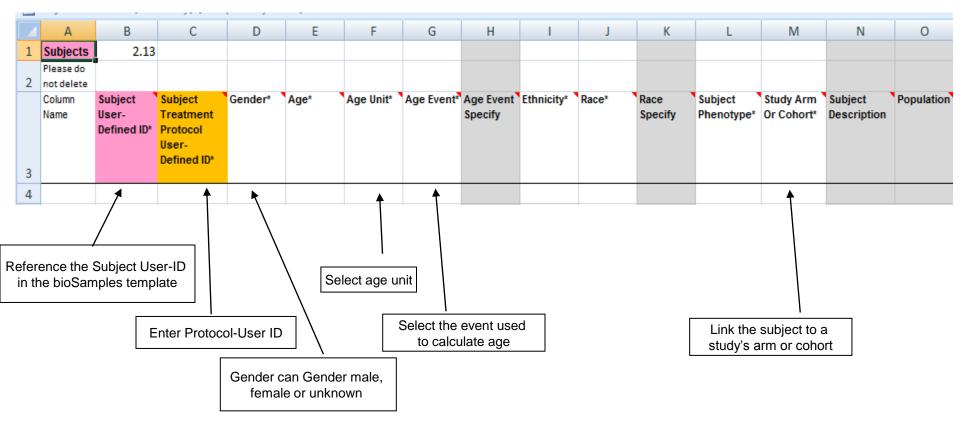




Subjects

Two templates for human (**subjectsHuman.xls**) and non-human primate subjects(**subjectsAnimal.xls**) are available to collect Subjects information. These templates are available from **Step 1 of Data Submission**. Only one subjects.xml file is available for both of these templates.

Note: Subjects MUST link to a Protocol(s). This is a requirement on DAIT minimum information standards.



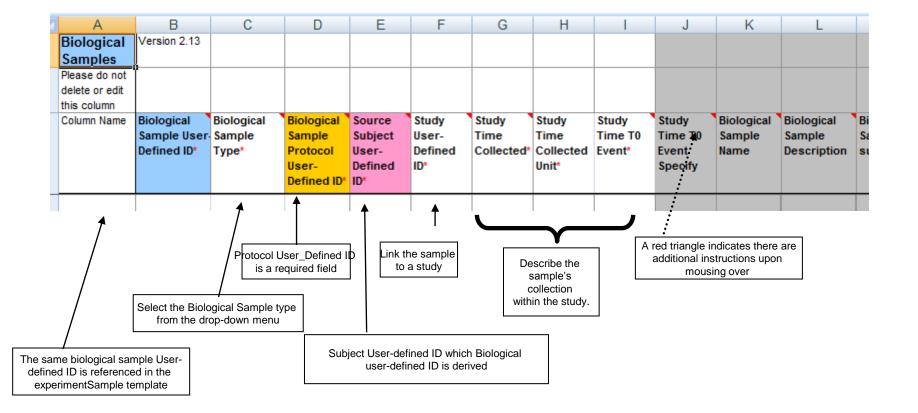


Biosamples

The **biosamples.xls** template captures biological material that has undergone isolation, processing and/or treatment prior to use in an experiment. The template is available from **Step 1: Submit Data**. Description on template may include links or references to a subject, protocols or a source biological sample.

bioSamples.xml is used to capture in a systematic way the metadata describing the biological samples used in an experiment. **Attention**: When an element uses a controlled vocabulary, the values are case sensitive and MUST be entered as specified.

Note: Biological Samples MUST link to a Protocol(s) and Subjects.

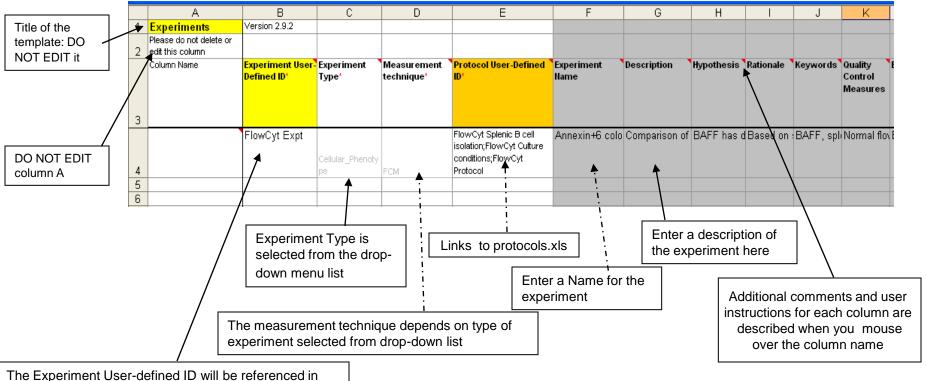




Experiments

The **experiments.xls** template captures the overall goal and methods of an experiment. The column "Experiment User-Defined ID" refers to the short name for the experiment. The "Experiment Name" column records a longer name. More detail can be provided about experiments, but that is the minimum data the user will need in filling out the template.

Note: Experiments MUST link to a Protocol(s). This is a requirement on DAIT minimum information standards.



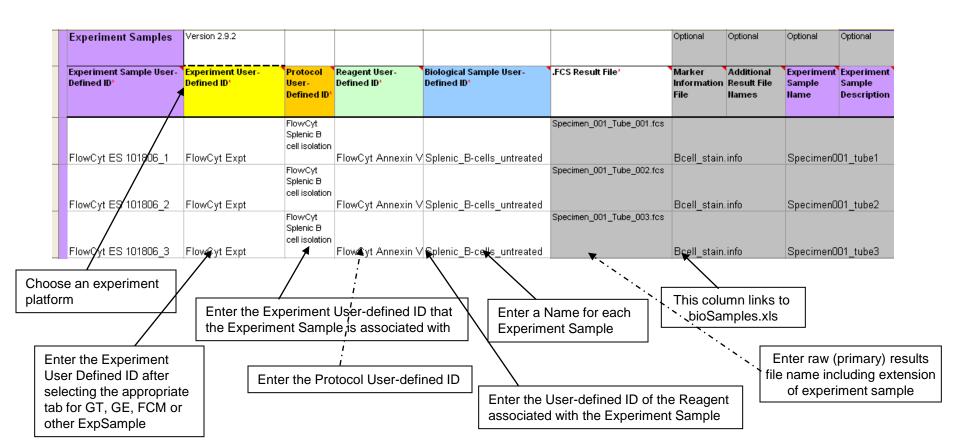
the experiment Oser-defined iD will be referenced in the experimentSamples.xls template to link together all experiment samples for an experiment. Each unique experiment should be entered in its own row



Experiment Samples

The experiment samples templates are organized by assay type. If there are commonly used file formats for an assay type (e.g. Affymetrix gene expression arrays or Illumina genotyping arrays), there is often an experiment sample template for the assay type and file format. The experiment sample templates all capture the links between experiments, protocols, biologial samples, and assay results.

When linking an experiment sample to more than one reagent, sample, protocol or result file, please use a semi-colon (;) to separate the records.



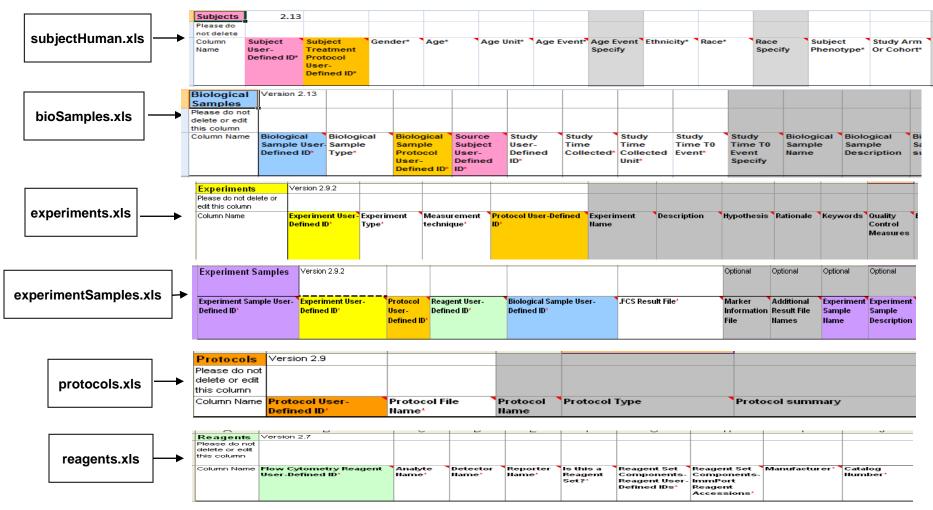
Experiment Samples templates are specific for the assay types below and reflect individual characteristics unique to each:

- ELISA
- ELISPOT
- Flow Cytometry
- Gene Expression
- Genotyping
- HLA
- KIR
- Mass Spectrometry
- MBAA
- qRT-PCR
- Hemagglutination Inhibition
- Virus Neutralization

For assistance with template completion, please contact the helpdesk at helpdesk@immport.org

Comparison of Templates

If the metadata files completed were stacked on top of each other, a new perspective on how each is related can be observed. The color coding of the metadata categories and how they are represented in different files can be seen.

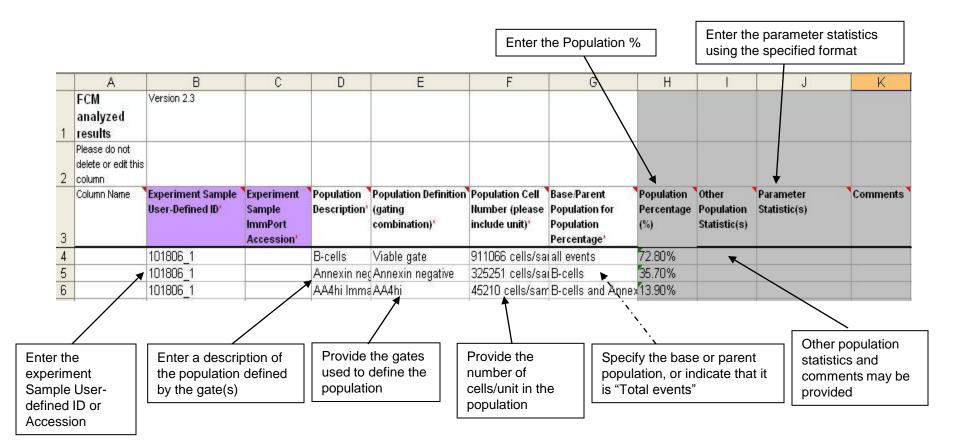




Derived Data

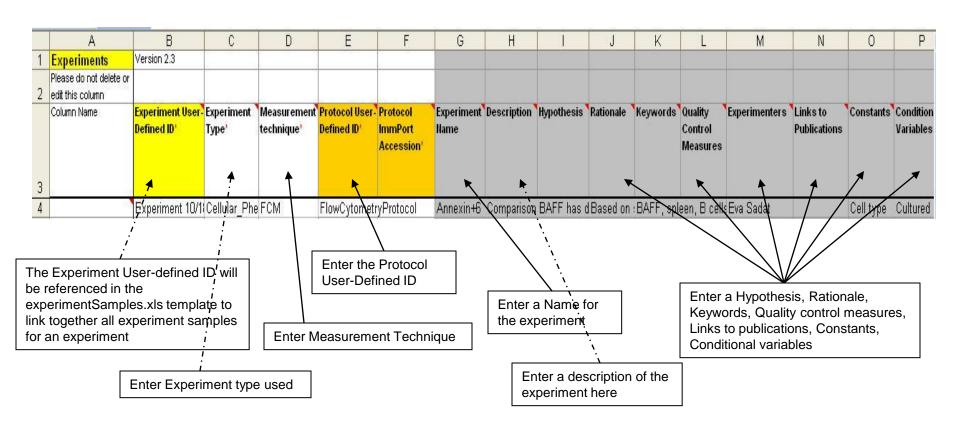
1. Flow Cytometry Derived Data Report

The **FCM_derived_data.xls** template is available from **Step 1 of Data Submission**. Flow Cytometry Derived data refers to the measurement of the Experiment Sample to which the population cell number, population percentage or other statistics apply



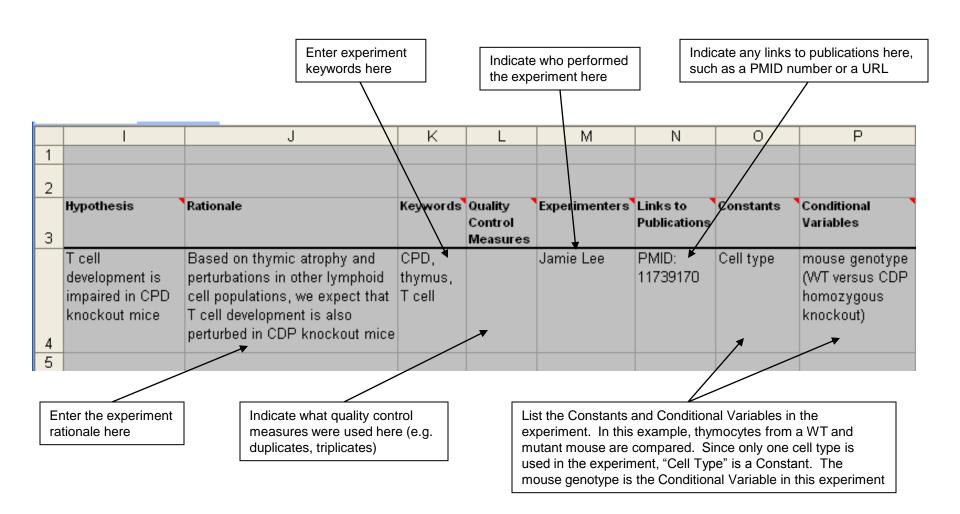


Example templates needed in Flow Cytometry data submission Experiments.xls template (1 of 2)





Example templates needed in Flow Cytometry data submission Experiments.xls template (2 of 2)





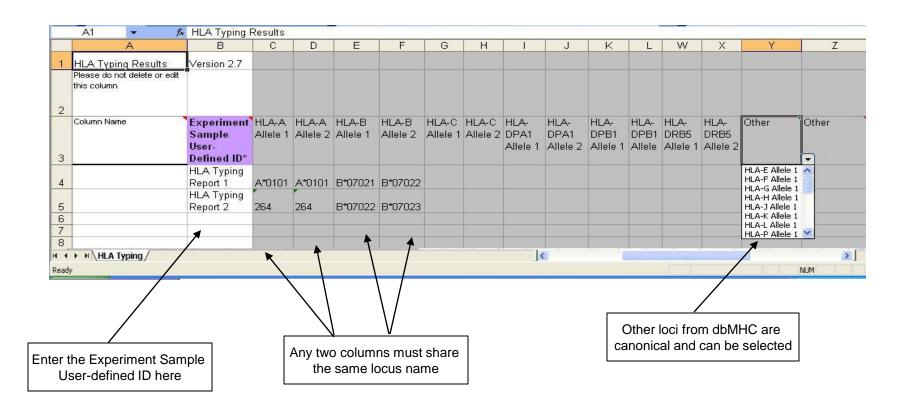
Derived Data, continued

HLA Typing Summary Report

The **HLA_Typing.xls** template is available at **Step 1 of Data Submission**. Data is used to analyze HLA using controlled vocabulary unique to groups of populations using measurements of gene types like HLA-A, HLA-B, DPA1, DPB1, DQA1, DQB1, DRB1etc. The loci names are available from the IMGT website (www.ebi.ac.uk/imgt/hla/) or the dbMHC (www.ncbi.lnm.nih.gov/mhc/) website. A horizontal format for collecting typing results is recommended.

Two columns must share a locus name with one column name ending in 'Allele 1' and the other column name ending in 'Allele 2'.

Note: Currently no validation of locus names is done.



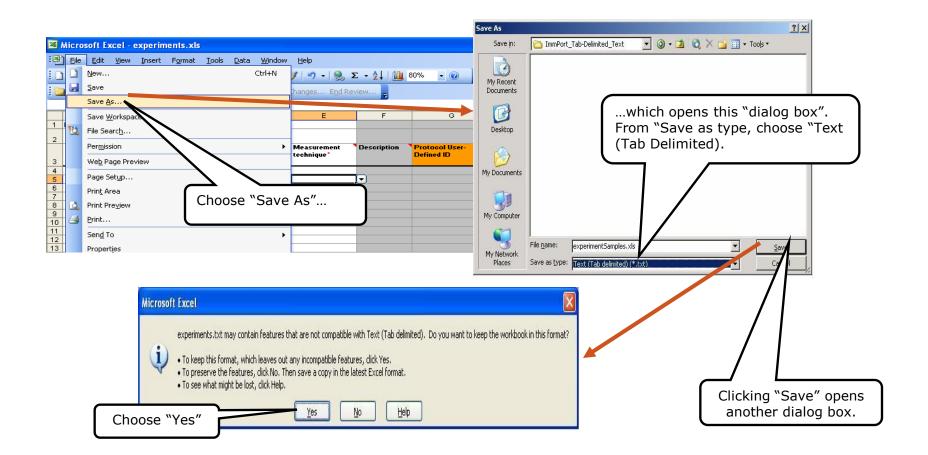


Save the spreadsheet file as tab-delimited text

Each template **must be saved in a tab-delimited text file (*.txt)** format in preparation for submission to ImmPort. Click on "File" in the Excel menu bar and choose "Save As". In the pop up window, click on the "Save as type" drop-down list, scroll down and choose "Text (tab-delimited) (*.txt)". When the pop window appears asking if the user wants to save the files as text, please select "Yes".

Note: The original names of the templates (e.g. protocols.txt) MUST be maintained.

Only one copy of each of the metadata files is included in a data submission package.

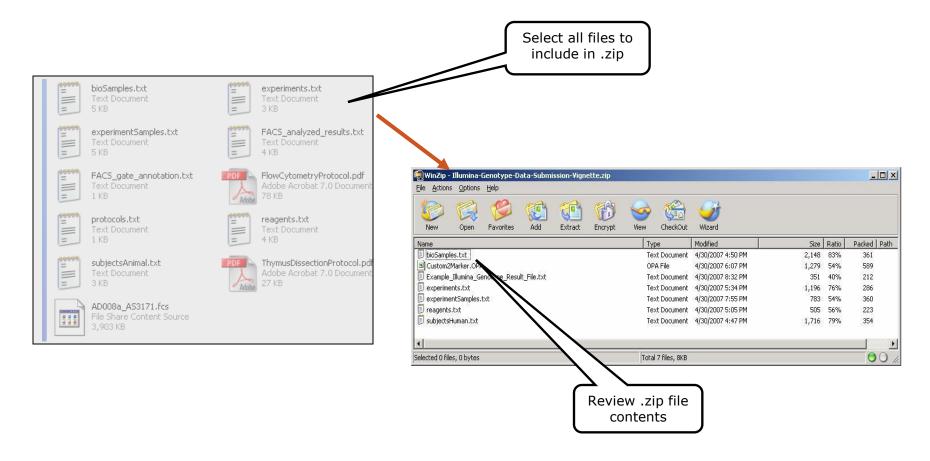


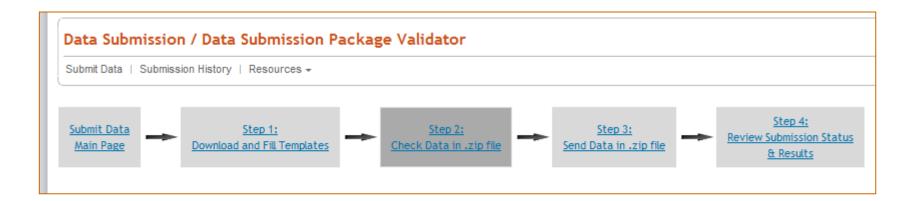


Data Submission Package .zip file contents and structure

The ImmPort data submission package is a .zip file that includes all of the files the user wants to submit. There are many tools that can be used to create a ".zip" archive file.

Select all of the files to be included and then create a .zip file (versus .zipping a folder with all of the files). Please do not select a folder or directory of files to ZIP- this confuses ImmPort's software.





Data Submission Package Validator

This tool checks the format and content of the files in a data package and reports issues. This tool is downloaded to your computer and run from a graphical user interface.

The Validator inspects completed templates (the metadata transfer files) for common formatting error and data integrity issues, checks for missing files, improperly formatted metadata files and references to other metadata within the data package, Integrity and structure of .zip files.

It is highly recommended to use the Validator before sending data into ImmPort as part of the data submission process.

When no issues are noted by the Validator, please submit the data submission .zip package to ImmPort.

The following requirements are important for the running the application:

- Java configured to run from the Windows command prompt.
- JRE 1.5.X or later installed.
- Minimum size of memory required is 64 MB.
- Number of files to be unzipped is 76 files and 42 folders.
- Disc space needed is 8.33 MB.
- No file overwriting will occur by default.



Uploading a file less than or equal to 1 Gigabyte -- Submit Online

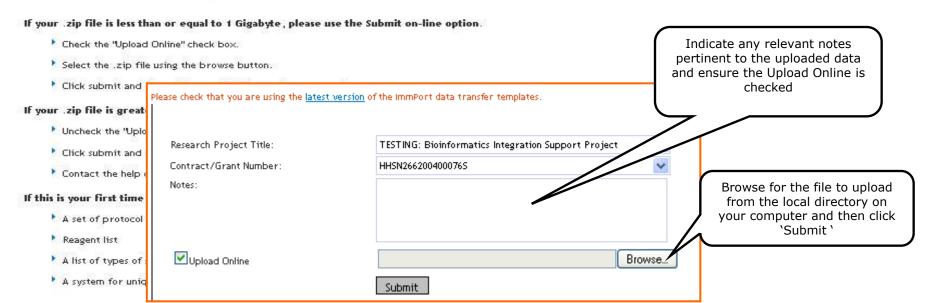
Click on the send data file tab and if the page has timed out, log back into ImmPort and continue with the submission process. Select the **research project** to which the data will be submitted. If none has been created, the Principal Investigator or the Project Manager can create one. If there are multiple projects listed, the Principal Investigator or the Project Manager decides which project to select.



Research data will be stored in the private project workspace of your research project. Please select the Research Project and Grant/Contract number with which your data will be associated and enter any comments in the Notes text box. If you need assistance in understanding what a Research Project is and whether you are associated with one, please contact the Help Desk.

ImmPort accepts data submission packages in the "ZIP" format. (Please see the Data Submission User Guide for more information on how to create a .zip file.)

Please DO NOT include spaces in the .zip file name.



Uploading a file greater than 1 Gigabyte -- Submit Offline

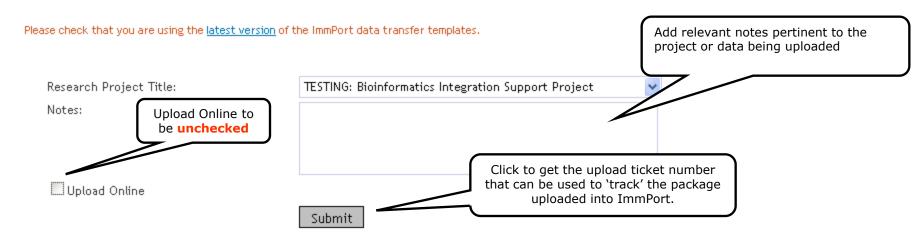
To upload data files greater than 1 Gigabyte uncheck the upload button and click submit. The Upload ticket is used to track the package for



Research data will be stored in the private project workspace of your research project. Please select the Research Project and Grant/Contract number with which your data will be associated and enter any comments in the Notes text box. If you need assistance in understanding what a Research Project is and whether you are associated with one, please contact the Help Desk.

ImmPort accepts data submission packages in the "ZIP" format. (Please see the Data Submission User Guide for more information on how to create a .zip file.)

Please DO NOT include spaces in the .zip file name.

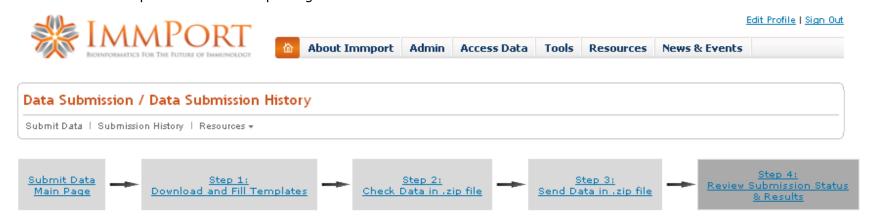




Review Data Submissions History

When the .zip file has been sent to ImmPort, the web page will be redirected to the Data Submission History page that lists the data packages that have been sent.

An email confirmation is sent to the submitter that the data package has completed submission processing or that there have been some problems with the package.



This page is a summary overview of the data submissions to a project. If you have access to more than one project, use the project filter to view the submission queue from other projects.

The data submitted to ImmPort can be queried and reviewed in Research Data Search

Filter By Project ADVN Varicella test Set Filter

Click on a Ticket Number below to view a submission's details.

15 items found, displaying all items.

The status of the submitted data is indicated as either 'Pending' or 'Completed'. The nature of the error is indicated on the Ticket Number for debugging

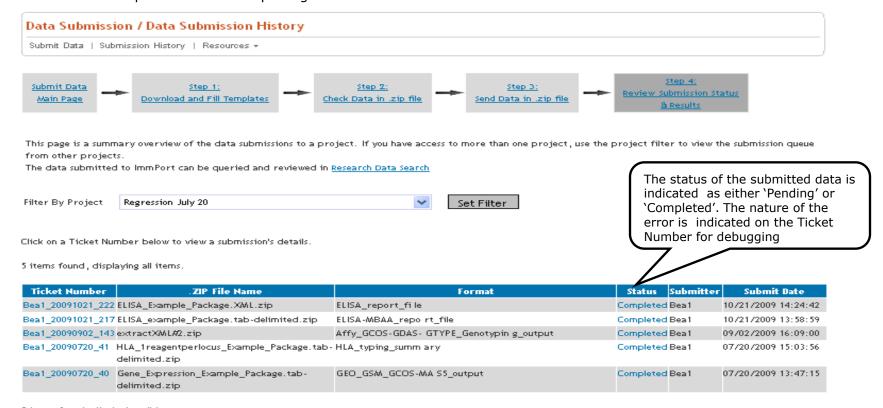
Ticket Number	.ZIP File Name	Format	Status	Submitter	Submit Date
kilelbe_20091013_194	ELISA_example_Package.tab-delimited.zip	ELISA-MBAA_repo rt_file	Completed	kilelbe	10/13/2009 09:21:11
kilelbe_20091013_195	ELISPOT_example_Package.tab-delimited.zip	ELISPOT_report_ file	Completed	kilelbe	10/13/2009 09:21:15
kilelbe_20091013_196	Flow_Cytometry_example_Package.tab-delimited.zip	FCM_report_file	Completed	kilelbe	10/13/2009 09:21:22
kilelbe_20091013_197	Gene_Expression_Example_Package.tab-delimited.zip	GEO_GSM_GCOS-MA S5_output	Started	kilelbe	10/13/2009 09:21:28



Review Data Submissions History

When the .zip file has been sent to ImmPort, the web page will be redirected to the Data Submission History page that lists the data packages that have been sent.

An email confirmation is sent to the submitter that the data package has completed submission processing or that there have been some problems with the package.



5 items found, displaying all items.

Legend:

- * Pending, The .zip file is in the data submission processing queue.
- * Started, The .zip file is being validated.
- * Completed, The .zip file is successfully processed and stored.
- * Rejected, One or more errors were encountered when processing the submitted files.